基于超图表示的服装兼容性预测模型 *

李 健 [1, 3†]，李 卓 [1, 2, 3]，马天祥 [4]，梁纪峰 [4]

(1. 天津大学 微电子学院，天津 300072; 2. 鹏城实验室，广东 深圳 200093; 3. 天津市成像与感知微电子技术重点实验室，天津 300072; 4.国家电网河北省电力有限公司电力科学技术研究院，石家庄 050021)

摘 要：针对现有服装兼容性模型都集中探究成对单品之间的兼容性这一问题，提出一种基于超图表示的服装兼容性预测模型。该模型首先基于现有数据集中时尚服装的不同类别和时尚服装间的搭配关系构建了一个服装超图，其中每个节点表示一件衣服，每条超边表示多件衣服组成的套装。为了更好地从超图中推断服装的兼容性，该模型将超图转换为传统图，并利用图神经网络模拟节点之间的复杂交互。最后引入注意力机制计算服装的兼容性得分，增强模型的预测能力。实验结果表明，在服装填空任务和服装兼容性预测两个服装搭配任务上，该模型分别达到了 77.29% 和 96.23%的准确率，较其他基线模型有显著的提升。

关键词：服装兼容性；超图表示；图神经网络；注意力机制

中图分类号：TP3 doi: 10.19734/j.issn.1001-3695.2022.01.0038

Outfit compatibility prediction model based on hypergraph representation

Li Jian[1, 3†], Li Zhuo[1, 2, 3], Ma Tianxiang[4], Liang Jifeng[4]

(1. *School of Microelectronics*, *Tianjin University*, *Tianjin* 300072, *China*; 2. *Peng Cheng Laboratory*, *Shenzhen Guangdong* 200093, *China*; 3. *Tianjin Microelectronics Technology Key Laboratory of Imaging & Perception*, *Tianjin* 300072, *China*; 4. *State Grid Hebei Electric Power Co., Ltd. Electric Power Research Institute*, *Shijiazhuang* 050021, *China*)

Abstract: In order to solve the problem of the existing research work focuses on the compatibility of paired items, this paper proposed an outfit compatibility prediction model based on hypergraph representation. The model constructs a fashion hypergraph based on the category information of fashion items and the collocation relationship between different fashion items in the existing dataset, where each hypernode represents an item, and each hyperedge represents an outfit made up of multiple items. To better infer outfit compatibility from the hypergraph, the model converts hypergraphs into traditional graphs, and the graph neural network is used to simulate the complex interaction between nodes. Finally, the attention mechanism is introduced to calculate the outfit compatibility score to strengthen the predictive ability of the model. Experimental results show that in the outfit fill-in-the-blank task and the outfit compatibility prediction task, the model achieved an accuracy rate of 77.29% and 96.23% respectively, which was significantly improved compared with other baseline models.

Key words: outfit compatibility; hypergraph representation; graph neural network; attention mechanism

0 引言

近年来，随着时尚行业的快速发展，服装在人们的日常生活中扮演者越来越重要的角色，一套得体的服装可以提升一个人的魅力并充分展示个性。据统计，在 2021 年的双十一购物节，阿里巴巴的电子商务网站天猫总交易额高达 5403 亿元，相比去年的 4982 亿元增加了 421 亿元。也就是说，这个潜在的市场有望创造巨大的财富并且服装的研究也因为这个巨大的市场而受到越来越多的关注。与此同时，服装相关的研究也随之兴起，例如个性化时尚设计[1]，服装组合[2]，单品推荐[3]和时尚趋势预测[4]等，尤其是服装搭配[5,6]方面的研究。然而，服装搭配是一项复杂的任务，它不仅取决于时尚风格、文化背景和潮流趋势等多个主观因素，所有的这些因素都可能因人而异，甚至随着时间的推移而变化，所以并不是每个人都能够搭配出合适得体的服装。因此如何对海量的服装进行分析，建立一种合适的服装搭配方法具有十分重要的社会意义和经济意义。

事实上，目前已经有很多工作致力于解决服装搭配的问题。比如，Viet 等人使用 SiameseNet 将套装中的单品服装从一个图像空间映射到一个风格空间，然后测量单品之间的距离来预测成对服装的兼容性[7]。Mcauley 等人提出使用 Low-rank Mahalanobis Transformation 将单品映射到样式空间来计算服装之间的相似性[8]。之后，Han 等人将一套套装表示为一组具有特定顺序的序列，并利用双向 LSTMs 来预测给定一组服装的下一件单品服装，以及组合套装的兼容性得分[9]。这些方法主要采用了两种服装表示方法：成对表示和序列表示。其中成对表示并不能反映多个服装之间的复杂关系，而对于序列表示来说，服装中并不存在固定的顺序。更重要的是，套装中物品之间的关系并不是有序的，因为每件物品不仅仅与序列中的前一件或者后一件物品有关系。而不管是基于成对表示还是序列表示的模型，它们考虑的都是成对服装之间的兼容性，这使得套装搭配兼容性最终的预测只依赖于成对单品之间的比较。然而，套装的兼容性不仅取决于成对服装之间的特性，还受同一套装中其他服装特征的影响。

因此，解决服装搭配兼容性问题的关键在于如何恰当地表示多件时尚单品之间的关系，而不是仅仅关注于成对的单品。近年来，一些研究方案通过图神经网络[10~12]解决服装搭配中的复杂关系问题。例如 Cui 等人利用类别信息将套装表

收稿日期：2022-01-28；**修回日期：**2022-03-28 **基金项目：**天津市自然科学基金资助项目(20JCQNJC01490)

作者简介：李健(1996-)，男(通信作者)，山西吕梁人，硕士研究生，主要研究方向为多媒体计算、机器学习(2019232064@tju.edu.cn)；李卓(1984-)，男，山西大同人，副教授，硕导，博士，主要研究方向为人工智能与数据智能化处理；马天祥(1986-)，男，辽宁朝阳人，高级工程师，硕士，主要研究方向为配网运行分析与配电自动化技术；梁纪峰(1985-)，男，高级工程师，硕士，主要研究方向为电力系统稳定分析与控制.

示为一个图，其中每个节点表示一个类别，每条边表示不同物品之间的不同关系，然后通过引入注意力机制输出套装兼容性得分[13]。之后 Cucurull 等人提出利用图自动编码器将服装兼容性问题看成是边预测问题，并且通过融合上下文信息提高服装搭配兼容性预测性能[14]。尽管使用图的各种模型在服装搭配问题上取得了成功的结果，但由于传统图的边是由两个节点相连而成，因此这种方法建立的也是成对节点之间的关系，并没有从本质上解决预测套装整体兼容性的问题。

为了解决上述的问题，本文提出通过超图来反映套装中多件单品之间的复杂关系和高层次关系。超图是一个可以表达复杂网络的广义概念。在传统的图中，边连接的节点数被严格定义为 2；而在超图中，每条边都可以连接两个以上的节点，这就使得每条边都可以表示一套完整的服装。为了更好预测套装搭配的兼容性，本文提出了一个新的模型 OCPCE 去模拟套装和时尚单品之间的交互关系。模型的框架被展示在图 1 中，该模型首先基于数据集构建了一个服装超图(即图 1 中最左边)，其中超图中的每个超节点代表不同的单品服装，每条超边代表多件单品组成的套装(样式相同的边为超边)；然后随机选取其中的一条超边进行兼容性预测。为了更好地表示超边和节点之间的复杂关系，模型将超边中的节点两两互相连接形成一个简单图，使得节点之间的交互可以更好的传播；之后在简单图上通过图卷积神经网络的消息传播机制聚合节点的邻居信息来迭代更新节点的状态信息表示；最后，在计算服装兼容性时，与现有工作认为所有单品对服装兼容性产生相同的影响不同，该模型引入注意力机制模拟不同单品对服装兼容性的影响，以更好地增强模型的预测能力。

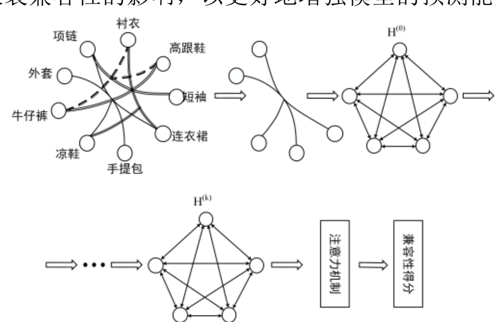


图 1　基于超图表示的服装兼容性预测模型框架图

Fig. 1　Framework of outfit compatibility prediction model based on hypergraph representation

1　基于超图表示的服装兼容性预模型

1.1　问题定义

服装搭配是否兼容需要考虑的是套装中所有时尚单品组合在一起之后的相容性，而不仅仅是成对单品之间的相容性。本文的研究目的是通过对套装和时尚单品之间的关系进行兼容性建模，从而预测服装整体的兼容性性。假设存在一个套装集合 $\mathcal{O}=\{o_1,o_2,\cdots,o_m\}$，其中 o_i 表示第 i 套套装，随机从集合中选择一套由多件单品搭配而成的套装，通过建模套装和时尚单品之间的关系，计算套装整体的兼容性得分并根据得分预测套装是否兼容。

1.2　特征提取和超图构建

a) 视觉特征提取　时尚单品的图片中包含了大量的信息，比如颜色、图案和条纹等等，对预测服装搭配兼容性有极大的帮助。本文利用卷积神经网络来提取时尚单品图像的视觉信息。相比于传统的特征提取方法如 SIFT、SURF 和 PCA 等，卷积神经网络已经被证明是图像特征提取中较先进的模型。本文选择由 Google 团队提供的经过预训练的 Inception-V3[15]深度神经网络进行视觉特征的提取。将时尚单品的图像输入到 InceptionV3 网络中，并将其线性层的输出作为该单品的视觉特征。每件单品服装的视觉特征维度是 2048 维。

b) 文本特征提取　时尚单品的文本信息主要来源于其自身的标题，标题中大多数文本均为单词或者较短的短语。因此本文通过词袋模型[16]来提取时尚单品的文本特征，该方法已经在文本特征提取[17]中被证明是有效的。首先基于数据集中不同时尚单品标题的单词构建一个词汇表。由于标题没有统一的标准规范，因此其中会出现很多没有意义的单词，比如 ‘a’、‘an’、‘de’ 等，因此需要过滤掉少于三个字符的单词以保证词汇的有效性。通过相关统计后得到一个包含 2757 个单词的词汇表，因此每件单品服装的文本特征维度是 2757 维。

c) 超图构建　本文基于现有数据集中时尚单品的类别信息和时尚单品之间的搭配关系构建了一个时尚超图 $\mathcal{H}=(\mathcal{V},\mathcal{E})$ 模拟套装和单品之间的复杂关系。其中，超图上的每个节点表示不同时尚单品所属的类别，每条超边(即样式相同的边)表示多个类别之间的搭配关系。但是并不是任意节点都可以构成超边。只有数据集中出现的套装中所包含的时尚单品类别所属的节点才可以构成超边。比如表示上衣、下衣、鞋子、包以及配饰等类别的节点可以构成超边。通过将每件单品服装输入到超图中对应的节点，超图中每条超边对应表示为一套完整的套装。

1.3　模型设计

本节首先介绍一些将在本节中使用到的符号。对于每件时尚单品 e_i，它对应的视觉特征是 s_i，文本特征是 t_i。每件单品 e_i 所对应的种类是 c_i，对应在超图中的超节点是 $v_i\in\mathcal{V}$，而超节点 v_i 在模型中的状态表示是 f_i。

a) 节点初始化　本文所提模型的输入是每件时尚单品的视觉特征 s_i 和文本特征 t_i，它们被用来初始化其相应节点的特征表示。对于每件单品，首先将它的视觉特征和文本特征映射到一个大小为 d 的空间。由于每件单品所属的种类都是不同的，因此在映射到样式空间的时候，每一个种类都需要设置一个不同的线性映射矩阵。然后拼接样式空间中的视觉特征和文本特征作为每件单品的特征表示。最后，初始化该特征表示作为每件时尚单品在超图中对应超节点的初始特征表示，如下所示。

$$\boldsymbol{f}_i^0=\tanh(\boldsymbol{W}_i s_i\,||\,\boldsymbol{W}_i t_i) \tag{1}$$

b) 超边转换为图　在本文所提模型 OCPCE 中，超图中的每条超边表示一套完整的套装。为了更好模拟套装和单品服装之间的复杂和高阶关系，该模型将将每条超边转换为传统图。目前将超边转为图的方式有两种，一种是连通分量拓展，即将超边中所有顶点都连接在一起，比如有三个顶点的超边，拓展成普通图时两两相连就会构成有三条边的简单图；另一种是星拓展，就是在每条超边中都增加一个新的节点，然后将超边中的所有节点与该节点相连接。第二种超边拓展方式在原来的超节点上增加了新的节点，这种方式可能会出现无效的信息，造成信息的错误传播。因此，本文通过连通分量拓展将超边转换为简单图。

c) 建模节点交互　超边转换为传统图以后，基于 GGNN[18]的方法，本文利用图神经网络建模传统图上的节点交互。图上的节点交互指的是每个节点聚合周围节点的状态信息和自身的状态信息来更新自身节点的状态信息。每次节点交互之后，节点的状态信息就更新为如下所示。

$$\boldsymbol{f}_u^{(k+1)}=\sum_{v_i,v_u\in\mathcal{V}}\boldsymbol{A}[v_i,v_u](\boldsymbol{W}\boldsymbol{h}_i^{(k)}+\boldsymbol{b}) \tag{2}$$

其中，$\boldsymbol{f}_u^{(k+1)}$ 表示的是节点 v_u 在第 k+1 次节点交互后的状态表示。$\boldsymbol{W}$ 和 $\boldsymbol{b}$ 是可训练的权值矩阵，用来提取有用的信息进行传播。$\boldsymbol{A}$ 是表示节点之间连通关系的邻接矩阵，如果节点 v_i 和节点 v_u 相互连接，则 $\boldsymbol{A}=1$，否则 $\boldsymbol{A}=0$。

在汇总邻居的状态信息后，每个节点通过 Gated Recurrent Units(GRU)来更新其最终表示，总体可以表示为

$$\boldsymbol{h}_u^{(k+1)} = GRU(\boldsymbol{h}_u^{(k)}, \boldsymbol{f}_u^{(k+1)}) \tag{3}$$

其中，$\boldsymbol{h}_u^{(k+1)}$ 为节点 v_u 在 k+1 次传播之后的最终表示。

d) 计算兼容性得分　为了评估多件衣服是否能构成一套匹配度高的套装，模型 OCPCE 通过每条超边中所包含节点的最终状态表示来计算超边所对应套装的兼容性得分。与现有工作中简单地将成对服装的兼容性聚合在一起作为套装的兼容性得分不同，本文认为不同的时尚单品对套装整体的重要性各不相同。因此，本文提出通过一种注意力机制来区分套装中各件服装的重要性。注意力机制的计算公式如下：

$$\hat{y}_u = \sigma(\boldsymbol{W}_1 \boldsymbol{h}_u^{(k)}) \tag{4}$$

$$\hat{x}_u = \delta(\boldsymbol{W}_2 \boldsymbol{h}_u^{(k)}) \tag{5}$$

其中，$\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$ 是两个可训练的权重矩阵，$\sigma(\cdot)$ 和 $\delta(\cdot)$ 分别是 LeakyReLU[19]和 Sigmoid 激活函数。$\hat{y}_u$ 表示不同单品对服装兼容性影响的重要性，$\hat{x}_u$ 表示套装中不同单品的兼容性得分。因此套装 o 的兼容性得分表示为

$$\hat{s}_o = \sum_{n=1}^{|m|} \hat{y}_u \hat{x}_u \tag{6}$$

其中，$|m|$表示套装 o 是由 m 件时尚单品组成的。

1.4　目标函数

为了更好的预测套装搭配的兼容性，本文采用 Bayesian Personalized Ranking(BPR)[20]算法来解决这个任务。在该算法中假设正样本套装比负样本套装有更高的兼容性得分。具体的目标函数表示如下：

$$\mathcal{L}_{bpr} = \sum_{(o,o^-)\in\mathcal{Z}} -\ln\eta(\hat{s}_o - \hat{s}_{o^-}) + \lambda\|\Theta\|_2^2 \tag{7}$$

其中，$\mathcal{Z}=\{(o,o^-)\}$ 是兼容性建模的数据集，每对 (o,o^-) 代表的是数据集中存在的套装 o (即正样本)和数据集中不存在的套装 o^- (即随机生成的负样本)。$\eta(\cdot)$ 是 Sigmoid 函数，Θ 表示所有的训练模型参数，λ 表示对其进行 L_2 归一化以避免过度拟合。

2　实验分析

2.1　数据集

现有的 Polyvore 数据集来自于流行的时尚网站 Polyvore.com，它允许他们的成员利用不同的服装创建时尚套装，或者喜欢并且保存其他人创建的套装。该数据集中包含 164379 件单品，形成 21899 套不同的服装。图分割技术被用来将数据集划分为训练集，测试集和验证集，其中 17316 套服装用于训练，1497 套用于验证，3076 套用于测试，并且划分出的数据集保证两两之间没有重叠，也就是测试集中出现的单品不会在训练集中出现。同时数据集中每个单品都包含着丰富的信息，比如图像信息，文本描述和类型(如牛仔裤、裙子、鞋子等等)。

如果原始数据集中套装的单品数超过了 8 件，表示该套装中包含重复出现的单品，而如果套装的单品数不足 3 件，表示该套装并不是完整的套装。因此本文通过移除原始数据集中超过 8 件单品和不足 3 件单品的套装生成新的数据集 Polyvore-N 来保持套装的不重复性和完整性。

2.2　实验设置

本文所有的实验都通过验证集对超参数进行选择，且所有的实验性能对比均在测试集上完成。为了优化目标函数，本文采用了随机梯度下降法，该方法已被证明在优化神经网络模型的过程中是有效的[21]。此外，本文采用网格搜索策略调整模型的超参数。在{8,12,16,20,24}范围中搜索批量的最佳大小，在$\{10^{-2},10^{-3},10^{-4},10^{-5}\}$范围内微调正则率和学习率，并且在{0,1,2,3}中搜索传播层数 k 来使模型达到最优效果。同时采用 Adam 优化器去优化整个预测模型和更新模型参数。所有实验均是在一台装有 Quadro M4000 图形处理器的服务器上进行的。实现训练过程在目标函数收敛或到达最大的循环次数时停止。

2.3　对比方法

Random：一个基于随机猜测的模型。

SiameseNet[7]：SiameseNet 送一对时尚单品进入到 Siamese 网络中，将它们映射到样式空间并比较它们之间的距离。通过平均成对单品的兼容性得分来计算整个套装的兼容性得分。

Bi-LSTM[9]：Bi-LSTM 采用双向 LSTM 挖掘一套服装中时尚单品之间的序列关系来对它们之间的兼容性进行建模，并计算它们之间的兼容性得分。

VCP[14]：VCP 方法引入了图自动编码器，根据两个时尚单品的视觉特征以及它们的上下文来计算它们之间的兼容性得分。

GGNN[18]：GGNN 利用图神经网络对套装和单品服装之间的关系进行建模，并计算它们之间的兼容性得分。

2.4　服装填空任务

服装填空任务(Fill In The Blank，FITB)是一项广泛开展的时尚兼容性研究的标准测试。给定一组搭配好的套装，将其中一件单品随机用空白代替，同时从数据集中任意选取三件单品作为错误选项，与被替代的单品一起组成候选集。本文将被空白代替的单品设置为正确答案，并且假设该单品比其他候选项与原套装更兼容。任务目的是从候选集中找到正确选项来填充套装中的空白。这项任务的性能通过从四个候选项中选择正确答案的准确率来评估。

本文提出的模型与其他可替代模型在 FITB 任务中的比较结果如表 1 所示。从该表中，可以得出以下结论：1)与其他方法相比，Random 的性能较差，表明仅仅通过随机猜测不足以反映套装整体的兼容性；2) Bi-LSTM 的性能优于 Random。原因可能是引入 Bi-LSTM 可以更好的学习有关兼容性的潜在知识。与随机猜测兼容性的模型相比，Bi-LSTM 将整个套装表示为一个序列，并在比较成对单品之外学习高阶关系；3)同样通过平均成对单品兼容性来计算套装兼容性的 VCP，性能却优于 Bi-LSTM。性能的提升要归功于 VCP 中引入了上下文信息来更好的反映单品之间的关系，这也证明了上下文信息在服装兼容性任务中的有效性；4)与其他方法相比，基于图的方法 GGNN 取得了更好的性能，表明图结构可以进一步有效地推断兼容性信息。与 Bi-LSTM 和 VCP 相比，这种性能的提升验证了图表示比序列表示和成对表示可以更好的对时尚单品之间的交互进行建模。5)本文提出的模型实现了最好的性能，这得益于超图表示可以很好的模拟套装和时尚单品之间的复杂关系和高阶关系。此外，该模型还引入自注意力机制来估计服装的兼容性，这可以更好的捕捉潜在的兼容性知识，进一步提高模型性能。如图 2 所示，本文随机选择几套示例服装可视化提出的模型在服装填空任务上的性能。

表 1　服装填空任务中不同模型的实验结果

Tab. 1　Experiment results of different models in outfit fill-in-the-blank task

模型	Accuracy (FITB)	模型	Accuracy (FITB)
Random	24.92%	GGNN	74.19%
Bi-LSTM	46.24%	OCPCE	77.29%
VCP	58.28%		

为了更加直观地分析所提模型 OCPCE 与对比模型 NGNN 以及 Bi-LSTM 在服装填空任务中的性能，本节从数据集中随机可视化了几个具体的样本案例，并对其进行详细说明。可视化结果如图 3 所示，其中方框表示模型选择正确的答案，而无方框表示模型选择错误的答案。在示例 1 中，

所有的模型都推断出示例套装缺少一双鞋子，并选择了正确答案。在示例 2 中，本文所提模型 OCPCE 和对比模型 NGNN 都选择了正确的答案，而 Bi-LSTM 模型却选择了错误的答案。这是因为 Bi-LSTM 模型是基于序列的方法，而上衣的位置与空白处相邻，会对空白处服装的选择产生极大的影响，最终导致 Bi-LSTM 选择了裤子作为答案。在示例 3 中，只有本文所提模型 OCPCE 选择了正确的时尚单品。虽然 NGNN 同样推断出了示例套装中所缺少的单品类别，但其选择的答案并不是与示例套装最兼容性的单品。而 Bi-LSTM 同样因为基于序列表示的原因选择了外套作为错误答案。本章所提模型 OCPCE 在三个示例套装中均选择了正确答案，说明基于超图表示的方法有效地模拟了套装和时尚单品之间的交互关系。上述示例验证了所提模型在服装填空任务上的优越性。



图 2 模型在服装填空任务中的可视化示例

Fig. 2 Visual example of model in the fill-in-the-blank task



图 3 不同模型在服装填空任务中的示例比较

Fig. 3 Example comparision of different model in fill-in-the-blank task

2.5 服装兼容性预测任务

服装搭配兼容性预测(Outfit Compatibility Prediction，CP)任务的目标是对给定的一套服装生成兼容性得分，该分数表示服装整体的兼容性程度。分数越接近 1 表示套装越兼容，越接近 0 表示套装越不兼容。这是现实生活中常见的问题，例如用户想要搭配一套适合自己的衣服，并希望确定时尚单品之间是否兼容。为了评估该模型，本文首先根据数据集构建兼容服装集，即数据集中的所有服装均为兼容服装。然后从兼容服装集中随机选择时尚项目，生成与兼容服装相同数量的不兼容服装集。通过对兼容服装集和不兼容服装集评分，来判断套装是否兼容。这项任务的性能通过采用广泛使用的 ROC(Receiver Operating Characteristic)曲线下的度量标准 AUC(Area Under Curve)来进行评估。

本文提出的模型与其他可替代模型在服装兼容性预测任务上的实验结果如表 2 所示。和 FITB 任务相似，本文提出的模型实现了最好的性能，表明超图的引入很好地揭示了时尚单品之间的高层次关系，增强了模型的预测性能。基于图表示的 GGNN 仍然也展现出了很好的性能，也反映出图结构仍然可以很好的结果服装兼容性任务。而引入上下文信息的 VCP 方法性能虽然优于序列表示 Bi-LSTM 和随机猜测 Random，但是他们在该任务中都没有展现出很好的竞争力，说明只对成对单品的兼容建模或者将套装表示为序列并不能很好的预测服装兼容性。从表中观察到的这些结果，同样支持了 FITB 任务中的分析。

表 2 服装兼容性预测任务中不同模型的实验结果

Tab. 2 Experiment results of different models in outfit compatibility prediction task

模型	AUC(CP)	模型	AUC(CP)
Random	50.12%	GGNN	94.77%
Bi-LSTM	77.11%	OCPCE	96.23%
VCP	90.13%		

2.6 不同组成模块对模型性能的影响

为了清晰地说明本文所提模型 OCPCE 在服装兼容性任务中的性能，本节从测试集中随机可视化了几套服装，并对其评分结果进行分析。如图 4 所示，OCPCE 模型为每套服装都进行了评分，其中得分越接近 1 说明套装兼容性越高，越接近于 0 说明套装兼容性越低。例如，第 1 套示例服装中各时尚单品之间种类互补并且颜色色调相近，因此它们获得了很高的兼容性得分；第 2 套服装中同时出现了两种不同样式的鞋子，这导致服装整体的兼容性得分由于单品种类重复出现而偏低；第 3 套示例服装中，虽然并没有重复出现相同类别的单品，但是可以很明显的发现套装中缺少了鞋子这一类别，因此也获得了较低的兼容性分数。综上，从示例中可总结出与常理相符的结论，套装中缺少或者重复出现某一类别的时尚单品都会使得服装的兼容性降低。同时，本文所提模型对这些案例的兼容性评分也验证了其在服装兼容性预测任务中的有效性。

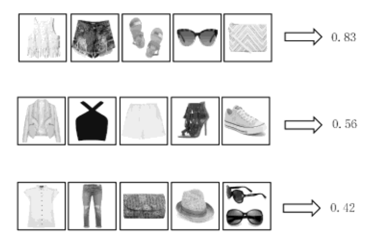


图 4 模型在服装兼容性预测任务中的可视化示例

Fig. 4 Visual example of model in outfit compatibility prediction task

2.7 不同组成模块对模型性能的影响

为了验证每个组成模块对提出模型性能的影响，本文通过禁用每个组成模块并比较它们的性能来进行消融研究。

实验结果如表 3 所示，其中 OCPCE(-h)表示该模型禁用超图模块后的变体，OCPCE(-w)表示禁用注意力机制模块后的变体，OCPCE(-h-w)表示同时禁用超图模块和注意力机制模块后的变体。从实验结果中可观察到以下结果：1)完整模型的性能优于所有的消融模型，证明了本文提出的模型中超图模块和注意力机制模块的重要性；2)禁用超图模块的性能比完整模型性能差，说明超图机制可以很好的模拟套装和时尚单品之间的高阶关系；3)禁用注意力机制的模型性能仅次

于优于完整模型的性能，这证明了注意力机制引入的必要性。

表 3 不同组成模块对模型性能的影响

Tab. 3 The effect of different components on model performance

模型	Accuracy (FITB)	AUC(CP)
OCPCE(-h-w)	75.28%	94.89%
OCPCE(-h)	75.94%	95.63%
OCPCE(-w)	76.07%	95.84%
OCPCE	**77.29%**	**96.23%**

2.8 不同模态对模型性能的影响

为了全面验证本文所提模型的有效性，本节在不同模态组合情况下对模型进行烧蚀实验。其中 OCPCE(VI)表示只有视觉模态、OCPCE(TE)表示只有文本模态与 OCPCE(VI+TE)表示文本模态与视觉模态相结合。

表 4 展示了本文所提模型 OCPCE 在不同模态下的性能比较。从表中可以看出：1)采用多模态信息的模型优于只采用单一模态的模型，说明文本模态和视觉模态都有助于提高兼容性建模的性能；2)只采用视觉模态的模型优于只采用文本模态的模型，说明影响时尚单品建模的服装因素更多地体现在视觉信息(如颜色和图案)，而不是文本信息(如材质和类别)。

表 4 不同模态对模型性能的影响

Tab. 4 Effects of different modality on model performance

模型	Accuracy (FITB)	AUC(CP)
OCPCE(TE)	75.78%	95.28%
OCPCE (VI)	76.34%	95.63%
OCPCE (TE+VI)	**77.29%**	**96.23%**

3 结束语

为了更好的预测服装搭配的兼容性，本文提出通过超图来表示套装和单品之间的关系，因为超图中的超边可以连接多个节点表示一套完整的套装。为了更好的从超图中推断套装是否兼容，本文将超边转换为传统图使得可以更好的捕捉套装中不同单品之间的复杂关系。同时使用真实数据集在不同类型的时尚搭配任务上进行实验，结果表明，本文所提模型可以有效的学习时尚服装的兼容性。但是由于每个用户都有独特的审美和穿搭风格，因此在之后的研究中，可以将用户的个人偏好融合到服装搭配技术中，通过衡量用户的体型、肤色等信息，实现基于用户的个性化搭配。

参考文献：

[1] Kang W C, Fang C, Wang Z, *et al.* Visually-aware fashion recommendation and design with generative image models [C]// Proc of IEEE International Conference on Data Mining. Piscataway, NJ: IEEE Press, 2017: 207-216.

[2] Feng Z, Yu Z, Yang Y, *et al.* Interpretable partitioned embedding for customized multi-item fashion outfit composition [C]// Proc of ACM on International Conference on Multimedia Retrieval. Piscataway, NJ: ACM Press, 2018: 143-151.

[3] Shih Y S, Chang K Y, Lin H T, *et al.* Compatibility family learning for item recommendation and generation [C]// Proc of the AAAI Conference on Artificial Intelligence. Piscataway, NJ: IEEE Press, 2018, 32 (1) .

[4] Al-Halah Z, Stiefelhagen R, Grauman K. Fashion forward: Forecasting visual style in fashion [C]// Proc of the IEEE international conference on computer vision. Piscataway, NJ: IEEE Press, 2017: 388-397.

[5] 刘锐，彭敦陆. 一种服饰风格特征指导下的服装搭配学习模型 [J/OL]. 小型微型计算机系统：1-6 [2021-12-23]. http://kns. cnki. net/kcms/detail/21. 1106. TP. 20210622. 1546. 008. html. (Liu Rui, Peng Dunlu. A fashion compatibility learning model guided by clothing style features [J/OL]. Journal of Chinese Computer Systems: 1-6 [2021-12-23]. http://kns. cnki. net/kcms/detail/21. 1106. TP. 20210622. 1546. 008. html.)

[6] 杨怡然，吴巧英. 智能化服装搭配推荐研究进展 [J]. 浙江理工大学学报：自然科学版，2021, 45 (01): 1-12. (Yang Yiran, Wu Qiaoying. Research progress of intelligent clothing matching recommendation [J]. Journal of Zhejiang Sci-Tech University: Natural Science, 2021, 45 (01): 1-12.)

[7] Veit A, Kovacs B, Bell S, *et al.* Learning visual clothing style with heterogeneous dyadic co-occurrences [C]// Proc of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2015: 4642-4650.

[8] McAuley J, Targett C, Shi Q, *et al.* Image-based recommendations on styles and substitutes [C]// Proc of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. Piscataway, NJ: ACM Press, 2015: 43-52.

[9] Han X, Wu Z, Jiang Y G, *et al.* Learning fashion compatibility with bidirectional lstms [C]// Proc of the 25th ACM International Conference on Multimedia. Piscataway, NJ: ACM Press, 2017: 1078-1086.

[10] Hong Richang, Li Lei, Cai Junjie, *et al.* Coherent semantic-visual indexing for large-scale image retrieval in the cloud [J]. IEEE Trans on Image Processing, 2017, 26 (9): 4128-4138.

[11] Hong Richang, Yang Yang, Wang Meng, *et al.* Learning visual semantic relationships for efficient visual retrieval [J]. IEEE Trans on Big Data, 2015, 1 (4): 152-161.

[12] Chen Long, Zhang Hanwang, Xiao Jun, *et al.* Counterfactual critic multi-agent training for scene graph generation [C]// Proc of IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2019: 4613-4623.

[13] Cui Zeyu, Li Zekun, Wu Shu, *et al.* Dressing as a whole: Outfit compatibility learning based on node-wise graph neural networks [C]// Proc of World Wide Web Conference. San Francisco: ACM Press, 2019: 307-317.

[14] Cucurull G, Taslakian P, Vazquez D. Context-aware visual compatibility prediction [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2019: 12617-12626.

[15] Szegedy C, Vanhoucke V, Ioffe S, *et al.* Rethinking the inception architecture for computer vision [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 2818-2826.

[16] Ji Rongrong, Xie Xing, Yao Hongxun, *et al.* Mining city landmarks from blogs by graph modeling [C]// Proc of the 17th ACM International Conference on Multimedia. Piscataway, NJ: ACM Press, 2009: 105-114.

[17] Gao Yue, Wang Meng, Zha Zhengjun, *et al.* Visual-textual joint relevance learning for tag-based social image search [J]. IEEE Trans on Image Processing, 2012, 22 (1): 363-376.

[18] Li Y, Tarlow D, Brockschmidt M, *et al.* Gated graph sequence neural networks [J]. IEEE Trans on Signal Processing, 2020, 68 (1): 6303-6318.

[19] Maas A L, Hannun A Y, Ng A Y. Rectifier nonlinearities improve neural network acoustic models [C]// Proc of the 30th International Conference on Machine Learning. Piscataway, NJ: IEEE Press, 2013: 1-5.

[20] Rendle S, Freudenthaler C, Gantner Z, *et al.* BPR: Bayesian personalized ranking from implicit feedback [C]// Proc of the 25th Conference on Uncertainty in Artificial Intelligence. Piscataway, NJ: ACM Press, 2013: 452-461.

[21] Zhang Tong. Solving large scale linear prediction problems using stochastic gradient descent algorithms [C]// Proc of the 21st International Conference on Machine learning. Piscataway, NJ: IEEE Press, 2004: 116.